Propensity Scoring for Comparison of Wound Care Treatments: A Case Study

by

Chenqi Fu

A thesis submitted to The Johns Hopkins University
in conformity with the requirements for the degree of
Master of Science and Engineering

Baltimore, Maryland
December, 2017

Abstract

Treatment effect modeling in an observational study is challenging. In an observational study, participants' conditions in diverse treatment groups may differ from each other to a large extent, which can result in potential bias in an analysis aimed at the comparison of treatment effect. The propensity score method is a commonly employed approach to analyzing causal effects and eliminating confounding based on an observation study. In this paper, two models for modeling propensity scores are discussed and implemented. Once models have been arrived at, treated subjects and control ones are matched according to estimated propensity scores. These methods are illustrated for an important application, where the effectiveness of two possible treatments for Diabetic Foot Ulcers are compared. Here, treatment effectiveness is measured that whether the wound is closed within 12 weeks after the treatment. At last, we conclude that there is no significant difference in the effectiveness of two treatments.

Acknowledgments

I would like to express my sincere gratitude to many people whose help was very valuable in this research.

I would first like to gratefully and sincerely thank my thesis advisor Professor Daniel Naiman of the department of Applied Mathematics and Statistics at Johns Hopkins University. The door to Prof. Naiman office was always open, especially when I had questions about my research or writing. His meaningful assistance, tireless guidance and patience are conducive to the accomplishment of this paper. He allowed this paper to be my own work, but steered me in the right direction.

I would like to thank Osiris Therapeutics, Inc. as well as the source of the database used in this work. Besides, the interesting visits to the company and inspiring lectures and discussions deepened my comprehension of the research. Particularly, I would like to thank Doctor Alla Danilkovitch for her continued support and encouragement. She has amazing knowledge of both wound care and statistics. Every time I had questions, her quick responses helped me out with queries and inspired me.

Table of Contents

Abstract ii

Acknowledgments iii

Table of Contents iv

List of Tables vii

List of Figures viii

1 Introduction 1

1.1 Diabetic Foot Ulcer 1

1.2 Propensity Score Matching 2

2 Methods 5

2.1 Notation and Assumptions 5

2.1.1 Notation 5

2.1.2 Assumption 7

2.2 Estimating Propensity Score 7

2.2.1 Binary Logistic Regression 8
2.2.2 Generalized Boosted Regression 8
2.3 Matching Methods 10
2.4 Postmatching Analysis 11
2.4.1 Computing Covariate Imbalance 11
2.4.2 Regression Adjustment Based on Matched Samples 12

3 Application **13**
3.1 Data Description 13
3.2 Variable Selection 15
3.3 Estimating Propensity Scores 17
3.3.1 Logistic Regression 17
3.3.2 Generalized Boosted Modeling 22
3.4 Matching and Covariate Balance 23

4 Results and Discussion **27**
4.1 Regression adjustment 27
4.2 Difference of Mean of Outcome Variables 27
4.3 Doubly Robust 28
4.4 Discussion 28

References **30**

A Appendix: R code **35**

Curriculum Vitae 41

List of Tables

3.1 Results for tests of univate association between variables and treatment . . 16

3.2 Results for logistic regression models fit relating for continuous variables to treatment. Covariates with p-value < 0.2 are regarded as potential associated. 17

3.3 Estimated coefficients of covariates using logistic regression. "*" in last column is significance code following the rule that "***" means p-value is in (0,0.001) and "*" means that p-value is in (0.01,0.05) 19

3.4 Summary of propensity score estimation using logistic regression. It shows the minimum, 1st quartile, median, mean, 3rd quartile, maximum and standard deviation of the propensity scores of two treatment groups. 20

3.5 Relative influence calculated by GBM 22

3.6 Summary of propensity score estimation using logistic regression 23

3.7 Matching sample size 24

3.8 ASAM for each covariates before and after matching based on estimated propensity score using logistic regression 24

3.9 ASAM for each covariates before and after matching based on estimated propensity score using GBM 26

List of Figures

3.1 Distribution of estimated propensity score using logistic regression. The x-axis shows the propensity scores of two groups and y-axis is the frequency of the propensity score. 20

3.2 Quantile-quantile plot of two treatments 21

3.3 Distribution of estimated propensity score using GBM 23

Chapter 1

Introduction

1.1 Diabetic Foot Ulcer

Ulceration of the foot caused by diabetes, one of the most severe and costly complications of Diabetes, is common nowadays. A diabetic foot ulcer (DFU) develops on the foot because of physical gravity during weight-bearing activity. Diabetic foot ulcers have a prevalence of 14% among diabetics. Disabling patients, diabetic foot ulcers frequently lead to amputation of the leg. It is estimated that 2 million Americans with diabetes are currently living with limb loss. Severe DFUs further contribute to high mortality rates. Another threat of diabetic foot ulcers is that healed wounds often recur. Despite treatment, ulcers readily become chronic wounds. The pathogenesis of foot ulceration, which is complex, exerts tremendous impacts on effectiveness of treatment for the immediate future and for the long-term effect of preventing recurrence. In 2009, the cost to U.S. hospitals for treating cases of DFUs was approximately $13 billion. Diabetic foot ulcers have been overlooked in health-care research for decades and clinical practices of wound care tends to be reliant on opinion more than scientific facts. In 2012, in the U.K., the treatment costs of DFU for the National Health Service approached £650 million, about £1 in every £150.

1.2 Propensity Score Matching

In fields such as psychology, economics and biostatistics, an observational study is often used to draw inference about causal effects from a sample to a population. Usually, observation studies provide information on real practice and give signals of benefits and risks. In contrast to controlled trials, independent variables in observational studies are not under control of research due to logistical constraints or experimental cost. Also, studies are carried out on data that are collected in a retrospective fashion. As a result, the most common challenge of observational studies is called *confounding*, which means that patients selected to be in a particular comparator group results in systematic differences between the profiles of the treatment groups, and probably leading to misunderstanding of causal effect of the outcome. Confounding of medical subjects, for instance, due to differing medications, potentially detrimental exposures, or interventions can arise from multiple sources. Usually, physicians prescribe medications and perform surgeries based on their prediction as to which treatments the patients are most likely to prove beneficial. This can result in biased conditions. Consequently, treatment groups may be tremendously imbalanced in terms of their covariate compositions.

Statisticians have developed a methodology of matching samples to reduce or even eliminate this imbalance. Supposing an observational study of which the goal is to compare treatment effect of two unique treatments, matching samples is applicable not only when independent variables are imbalanced, but when there are more control samples than treated. The typical idea of matching is to identify a subset of control groups that has similar characteristic of covariates in the treated groups so that researchers reduce imbalance in the covariates' empirical distribution. Here, similarity of patients for the purpose of matching can be measured using Mahalanobis distance between their covariates as proposed

by Cochran and Rubin (1973) and Rubin (1974, 1978a). However, analyses of matched sampling based on Mahalanobis distance becomes increasingly infeasible when the number of covariates is increased, due to the fact that the capability to find appropriate subset of control groups for treated samples decreases. It is also the case that categorical variables may require special treatment.

Propensity score, introduced by Rosenbaum and Rubin (1983), is an alternative characteristic that can be used for matching. In fact, another difficulty of observational studies, in addition to covariates imbalance, is that independent variables or other features of subjects may be related to both treatment selection and experimental outcomes. A propensity score defined for each subject in the study is the conditional probability to receive particular one of candidate treatments given the covariates. Propensity scoring is useful as an advanced technique for matching. On the one hand, the propensity score has been proven by Rosenbaum and Rubin (1983) to be a balancing score in the sense that the conditioned on the propensity score, treatment and covariates become independent. On the other hand , the use of the propensity scoring addresses the problem of dimensionality that Mahalanobis distance has because it is univariate score. As a consequence, analysis is carried out by, across treatment, comparing subjects possessing similar propensity scores.

However, in reality, propensity scores are unknown for most of studies, and so statisticians must fit models to estimate them. To remedy this, it is common practice to fit a logistic regression model in that most of studies compare two treatments leading to the binary variable indicating the treatment selection. Breiman (2001) suggested that many classification algorithms from the machine learning literature can be used for modeling propensity scores, such as neural networks, linear classifiers, decision trees and meta-classifiers.

After estimation, there are numerous applications of propensity scores, including matching, subclassification and weighting. In health care research, propensity scores are useful when observational studies are quite common.

This paper proceeds as follows. Section 2 presents two models to estimate propensity scores and introduces methods to implement matching based on estimated propensity scores. Section 3 applys the models and a matching method in a case study of DFU treatment. We conclude with a brief discussion in Section 4.

Chapter 2

Methods

2.1 Notation and Assumptions

2.1.1 Notation

We consider a situation when a patient can receive exactly one of two possible treatments. Inference about efficacy of treatments on an individual involves the speculation about how this treatment would have performed on the patient if the treatment not received had been assigned. In a case of a treatment effect comparison, we use the notation for the treatment indicator $z_i = 1$ if the ith individual receives the treatment under investigation and $z_i = 0$ if he/she receives the placebo. Consequently, the potential outcomes of individual i are denoted by r_{i1} and r_{i0} where $i = 1, 2, \ldots, N$ and N is the sample size. The treatment effect for an individual can be written as

$$\tau_i = r_{i1} - r_{i0}.$$

To compare two treatment effects, parameters of interest are "average treatment effect on the treated"(ATT) and "average treatment effect"(ATE). Literally, ATT is defined as

$$\tau_{ATT} = E(\tau|z = 1)$$

$$= E(r_{i1}|z = 1) - E(r_{i0}|z = 1).$$

And ATE is defined as

$$\tau_{ATE} = E(r_1) - E(r_0).$$

In randomized controlled trials, which is rare in social sciences and natural sciences cases, the treatment assignment and covariates are independent.Thus, we note that $\tau_{ATE} = \tau_{ATT}$. In observational study, in general, there is a difference.

One possible estimation strategy for treatment effects which is actually widely used to solve the problem of speculation of counterfactual outcomes, suggested by Rosenbaum and Rubin (1983) is the propensity score matching method.

The propensity score is the conditional probability of a patient being assigned to a particular treatment given observed covariates X. That is $b(x) = Pr(z = 1|X = x)$. Rosenbaum and Rubin (1983) show that the propensity score is a balancing score given which the average treatment effect and treatment are conditionally independent

$$x \perp z|b(x).$$

Then for all x the average treatment effect at $b(x)$ is equal to the expected difference in observed response at $b(x)$

$$E(r_1 - r_0|b(x)) = E(r_1|b(x), z = 1) - E(r_0|b(x), z = 0).$$

Consequently, the ATT could be rewritten as

$$\begin{aligned}\tau_{ATT} &= E(r_{i1}|z=1) - E(r_{i0}|z=1) \\ &= E_{b(x)|z=1}(E(r_1|z=1,b(x)) - E(r_0|z=1,b(x))) \\ &= E_{b(x)|z=1}(E(r_1|z=1,b(x)) - E(r_0|z=0,b(x))) \\ &= E_{b(x)|z=1}(E(r_1 - r_0|b(x))).\end{aligned} \tag{2.1}$$

In Section 4, we conclude treatment effects of two treatments using ATE.

2.1.2 Assumption

There are two key assumptions for propensity score matching.

First, we assume participation is independent of outcome conditional on covariate, which can be expressed as:

$$E(\gamma_0|X, z=1) = E(\gamma_1|X, z=0)$$

This assumption is also called *ignorable treatment assignment assumption* (Rosenbaum & Rubin, 1983)

The second assumption is the *stable unit treatment value assumption* (SUTVA). SUTVA is the priori assumption that the treatment effects of a specific patient when assigned to a specific treatment will be the same whatever assignment mechanism is used.

2.2 Estimating Propensity Score

There are several methods for estimating propensity score using a vector of observed covariates. These methods includes logistic regression, the probit model and generalized

boosted model which is based on regression trees.

2.2.1 Binary Logistic Regression

Most of the analyses based on propensity scores make use of logistic regressions to estimate propensity scores of subjects. That is

$$\log \frac{b(X)}{1-b(X)} = \alpha + X^T\beta, \tag{2.2}$$

where α and β are logistic regression coefficients at which the mean square error is minimized. Plugging these estimated parameters in equation 2.2 and solving for b(x) to give

$$b(x) = \frac{1}{1+e^{-(\hat{\alpha}+X^T\hat{\beta})}}$$

we obtain the predicted propensity score for each sample patient i

The coefficients are estimated by maximum likelihood.

Logistic regression is feasible as well as attractive for two reasons. On the one hand, it naturally produces probabilities in the range $(0,1)$. On the other hand, logistic regression is a well-developed tool that can be readily implemented in most of statistical packages. We will make use of this model assumption in Section 3.

2.2.2 Generalized Boosted Regression

Logistic regression sometimes may not be a good of fit in that it requires of the log odds ratio in the covariates, which is not a general situation of case studies. Beside, logistic regression requires observations to be independent, then assumes linearity of independent variables and log odds. Since logistic regression is specifying an unknown functional form, there is an

increasing demand for better alternatives. Hence, there are several algorithm from machine learning studies that can be applied to propensity score estimation, including generalized boosted modeling (GBM).

Ridgeway (1999), Friedman (2002), and Mease, Wyner, and Buja (2007) proposed the GBM approach as the latest prediction methods. GBM is a general, data-adaptive algorithm that fits several models by the way of classification and regression trees. The model is comprised of many simple regression trees build iteratively. To begin with, the first decision tree is the one that minimizes the loss function. At the second step, a tree is fitted to the residual of the first tree which is allowed to contain different variables and split differently, still minimizing the loss function. The final GBM model is a combination of many trees that can be regarded as a regression model. The algorithm stops when it reaches the minimum of average standardized absolute mean difference(ASAM) of covariates. The ASAM is defined by

$$ASAM = \frac{1}{p} \sum_{covariates} \frac{|mean(treated) - mean(control)|}{SD(treat)},$$

where p is the number of covariate.

Because GBM does not specifying a functional form, it does not produce regression coefficients as logistic regression. Instead, it introduce relative influence, which is the percentage of log-likelihood explained by each variable. Suppose that using generalized boosted modeling, the approximation to the true model is $\hat{F}(\mathbf{x})$. Friedman (2001) suggested that one measure of relative influence I_j, of input variable x_j, is defined as:

$$I_j = (E_x[\frac{\partial \hat{F}(\mathbf{x})}{\partial x_j}] \cdot Var_x[x_j])^{1/2} \tag{2.3}$$

One can use package *gbm* in R to estimate propensity scores.

2.3 Matching Methods

The motivation behind propensity score matching is that groups of individuals with the same propensity score can be analyzed in a simple manner. There are various methods available for propensity score matching.

The most straightforward method is nearest neighbor. Here, the method involves randomly choosing an individual from the treated group then finding the counterpart of it in control group in terms of the closest propensity score. However, there are two approaches to identify those counterparts, with or without replacement. When we consider matching pairs without replacement, analyses produce different matched results dependent on the order in which observations get matched.

An intuitive extension of nearest neighbor matching is to match the treated and the control by subgroups. In a more common implementation of propensity score matching, pairs of treated and control subjects are formed whose propensity scores differ by a pre-specified amount (the so-called caliper width). Within a certain caliper width, an individual can be matched to the a set of counterparts. Moreover, a group of individuals within a specific range of propensity scores can be matched to corresponding control group within the same interval of propensity score.

It is common that there are more control samples than treated. However, instead of discarding unmatched samples from the control group, a kernel weight could be used to estimate the counterfactual treatment effect. In kernel matching, a weighted composite of comparison observation is collected to give the match for each treated patient, while the weights are evaluated by their distance in terms of propensity score from the treated.

There is no clear optimal method for matching. A standard recommendation is to test

several methods and choose the one that give rise to the best balance of the samples.

2.4 Postmatching Analysis

2.4.1 Computing Covariate Imbalance

Haviland et al. (2007) suggested a measure for covariate imbalance called *absolute standardized difference in covariance mean*, which is similar to ASAM in literature.

We use d_X to describe covariate imbalance before matching, and d_{X_m} after matching. There are calculated using the following formula:

$$d_X = \frac{|M_{Xt} - M_{Xp}|}{S_X} \tag{2.4}$$

$$d_{X_m} = \frac{|M_{Xt} - M_{Xc}|}{S_{X_m}} \tag{2.5}$$

$$S_X = \sqrt{\frac{(S_{Xt}^2 + S_{Xp}^2)}{2}} \tag{2.6}$$

$$S_{X_m} = \sqrt{\frac{(S_{Xt}^2 + S_{Xc}^2)}{2}} \tag{2.7}$$

where M_{Xt}, M_{Xp} and M_{Xc} are the means of a variable of treated group, potential control group and control group matched to treated participants, respectively. S_{Xt}, S_{Xp} and S_{Xc} are the standard deviations of treated group, potential control group and matched control group.

2.4.2 Regression Adjustment Based on Matched Samples

After acquiring matched samples using matching methods, Rubin (1978b) proposed that one can estimate ATE by using a specific type of regression adjustment. The regression adjustment is applied as follow:

1. Take the difference of the outcome variables between treated and control groups, $Y = Y_1 - Y_0$.
2. Take the difference of the covariates between treated and control groups, $X = X_1 - X_0$.
3. Fit regression model of Y on X which is $Y = \hat{\alpha} + X\hat{\beta}$.

Then $\hat{\alpha}$ is the ATE. We can use t-statistic and p-value of regression coefficient to perform a hypothesis test.

Chapter 3

Application

3.1 Data Description

The data chosen for this method in the application is collected from an electronic health record (EHR) database (NetHealth, Pittsburgh, PA). For the purpose of analyzing the treatment effect of the target new treatment compared with the prevailing one, all treatment records of patients who suffered from DFUs (Diabetic Foot Ulcers) with records from July 1, 2012 through June 30, 2016 were extracted from the database.

The database includes a variety of files providing diverse information about each patient and the treatment, such as the visit record, medical history, wound assessment and some personal information. In total, there are $4,588$ patients who have wounds on their feet and legs, including not only diabetic ulcers but other types of wounds, such as venous ulcers, trauma wounds and burns. Each patient may have more than one wound so that we have $34,879$ wounds recorded. Among those wounds, there are $10,264$ diabetic ulcers and $8,683$ of them are DFUs.

The primitive treatment records included patient's baseline demographics, wound location, size and duration, and wound-specific information recorded at each visit, including

size measurements and treatments. Wound measurements of length and width are used to calculate wound area in cm^2.

In the course of data processing, it is those patients who have DFUs located on foot, toe, heel, metatarsal head, toe web space, toe amputation site, or trans metatarsal amputation site that are of the most interest for scientific purpose.

Additionally, since we focus on the comparison between the new treatment and the reference treatment, those patients who did not receive either of them are eliminated. After omitting the patients with missing value of area of wound at time of treatment and the treatment duration, there are 1,594 samples and 282 covariates, including 100 covariates recording the area of wound at every day, if it is available, after the first assessment.

Several covariates are special because they can be used to model the propensity score. A patient attribute that is measured up to the time of treatment is referred as a pre-treatment covariate. Such covariates could exert influence on the choice of treatment. The pre-treatment covariates of interest are age, gender of the patient, smoking status and exposure of the body tissue, such as tendon, ligament, joint or bone, which could be measures of the severity of the wound. These are either binary or are categorical variables having either 4 to 5 groups. Additionally, since the DFU is superinduced by diabetes, whether it depends on the type of the diabetes is of concern. Wound length and width that we collected at the beginning of the treatment are quite likely to be related to both treatment effects and treatment selection. These variables not only contribute to extent of the individual treatment effect, but the option of which treatment to apply may largely rely on them.

Other covariates, for instance, the number of wounds, the number of visits, and days after surgery differ from patient to patient. And some variables that are available to use are not included in the analysis because a priori they are not viewed as relevant. These

variables includes the state in which a patient resides, the wound center where the patient was treated, and medical history beyond diabetic status.

3.2 Variable Selection

At the first stage of fitting the propensity score, a logistic regression model is highly recommended. To decide what covariates should be used to fit the logistic model, we evaluate the univariate association of each single covariate with the treatment. For categorical variables, we introduce contingency tables to examine the relationship and eliminate those with p-value higher than 0.2 based on a χ^2 test. For continuous variables, a univariate logistic regression model is fitted and the variable is included if a test of its significance is less than 0.2.

Patient age is regarded as an indicator of the physiologic stage of the patient. We analyzed age both as a continuous variable and by breaking it up into categories as "younger", "young", "middle age", "old" and "elder" using ventiles as break points. The "younger" patients are younger than 54 years old. Patients who are elder than 54 years old but younger than 61 are in category "Young". Patients who are elder than 61 years old but younger than 68 are in category "Middle age". Patients who are elder than 68 years old but younger than 75 are in category "old". Patients who are elder than 75 are in category "elder". After applying contingency tables, p-value > 0.5 shows that age group is not influential for choice of treatment. And even applying logistic regression, there is no evidence that age is crucial component for the decision of treatment.

Specifically, all categorical variables considered are "body.part", "smoking", "gender", "exp.tendon", "exp.ligament", "exp.muscle", "exp.joint", "exp.bone", "exposed", "dorsal.plantar", "proximal.distal", "diab.type", "hypertension".

"Smoking" describes the smoking habits of the patients as "current" for regularly smoker, "former" for who is used to smoke but a long period elapsed since last strike, "never" for non-smoker and "Unknown" for missing the record of smoking status of that patient.

And other covariates are binary variables with 1 means Yes, 0 means No.

	covariates	X_squared	DF	p_value	Signif
1	body.part	26.4882	21	0.1884	*
2	smoking	7.6403	3	0.0541	*
3	gender	3.4875	1	0.0618	*
4	exp.tendon	9.2178	2	0.0100	*
5	exp.ligament	3.2668	2	0.1953	*
6	exp.muscle	20.9086	2	0.0000	*
7	exp.joint	2.2620	2	0.3227	
8	exp.bone	8.8282	2	0.0121	*
9	exposed	15.6322	1	0.0001	*
10	dorsal.plantar	7.2913	2	0.0261	*
11	proximal.distal	2.0266	2	0.3630	
12	diab.type	3.0462	1	0.0809	*
13	hypertension	102.1002	91	0.2003	

Table 3.1: Results for tests of univate association between variables and treatment

Table 3.1 shows the results for testing univariate associations between categorical variables and treatment by using contingency tables. Variables with a star marked in the column Signif represents they are considered to be potentially associated with treatment selection.

From table 3.1, we conclude that variables with the p-value of contingency tables lower than 0.2 are "body.part", "smoking", "gender", "exp.tendon", "exp.ligament", "exp.muscle", "exp.bone", "exposed", "dorsal.plantar" and "diab.type".

Table 3.2 shows the result of fitting logistic regression to each continuous variable, including estimated coefficients, zvalue and p-value.

The continuous variables are "age", "bmi", "length", "width", "area", "depth", "ndfu",

"nwounds", "nsurgwounds". "age" is the patient's age at the first visit to wound center. "length", "width" and "depth" describe the length, width and depth of the wound when the wound was treated for the first time. "nwounds" indicates the total number of wounds the patient had. "nsurgwounds" shows how many wounds were treated. In addition, we have "length", "width", "area","depth" and "nsurgwounds" which may be of potential significance in choosing from treated options.

	covariates	Estimate	Std.Error	z_value	p_value	Signif
1	age	0.0027	0.0044	0.6087	0.5427	
2	bmi	-0.0086	0.0069	-1.2399	0.2150	
3	length	0.0109	0.0030	3.6367	0.0003	*
4	width	0.0142	0.0040	3.5861	0.0003	*
5	area	0.0171	0.0057	3.0157	0.0026	*
6	depth	0.0545	0.0148	3.6732	0.0002	*
7	ndfu	0.0064	0.0116	0.5503	0.5821	
8	nwounds	0.0046	0.0067	0.6936	0.4879	
9	nsurgwounds	0.0795	0.0562	1.4142	0.1573	*

Table 3.2: Results for logistic regression models fit relating for continuous variables to treatment. Covariates with p-value < 0.2 are regarded as potential associated.

3.3 Estimating Propensity Scores

3.3.1 Logistic Regression

Using the R package MASS, we use the stepwise bidirectional elimination approach for model selection. Bidirectional elimination is a combination of forward and backward approaches, testing at each step for variables to be included or excluded. This method uses Akaike information criterion (AIC) at each step to decide on the addition or deletion of a

covariate from consideration. We get the smallest AIC formula as

$$\begin{aligned} treatment \sim & smoking + gender + exp.muscle + dorsal.plantar \\ & + diab.type + width + depth + nsurgwounds \end{aligned} \tag{3.1}$$

Then, we consider the covariate interaction terms. We add all interaction terms, for instance, "smoking × gender", "smoking × depth" and etc. into the model. Then we also use the stepwise variable selection approach to find the best model which has the smallest AIC:

$$\begin{aligned} treatment \sim & \ smoking + gender + exp.muscle + dorsal.plantar \\ & + diab.type + width + depth + nsurgwounds \\ & + smoking : depth + smoking : nsurgwounds + gender : exp.muscle \\ & + gender : diab.type + gender : nsurgwounds + exp.muscle : depth \\ & + dorsal.plantar : depth + diab.type : nsurgwounds \end{aligned}$$

Table3.3 shows the estimated coefficients in the model. We can see that "plantar" and the interaction of unknown smoking status and the number of treated wounds are of high significance, while the width, the number of treated wounds, the interaction term "smokingFormer × depth", "genderM × nsurgwounds" and "dorsal.plantarPlantar × depth" are of significance. Especially, when patients have wounds on plantar and the number of treated wounds are large, they are more likely to receive treatment A.

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.9651	1.3367	0.72	0.4703	
smokingFormer	0.4857	0.4120	1.18	0.2385	
smokingNever	0.1725	0.4445	0.39	0.6980	
smokingUnkown	0.6428	0.4291	1.50	0.1341	
genderM	-1.3074	1.1703	-1.12	0.2639	
exp.muscleYes	-1.0978	0.6940	-1.58	0.1137	
dorsal.plantarPlantar	1.1269	0.3325	3.39	0.0007	***
dorsal.plantarUnknown	0.6248	0.3179	1.97	0.0494	
diab.type2	-0.9287	1.2397	-0.75	0.4538	
width	-0.0090	0.0046	-1.98	0.0482	*
depth	0.1183	0.0835	1.42	0.1567	
nsurgwounds	1.3310	0.5986	2.22	0.0262	*
smokingFormer:depth	-0.1121	0.0556	-2.02	0.0438	*
smokingNever:depth	0.0277	0.0605	0.46	0.6468	
smokingUnkown:depth	-0.0208	0.0592	-0.35	0.7255	
smokingFormer:nsurgwounds	-0.2421	0.1878	-1.29	0.1975	
smokingNever:nsurgwounds	-0.3191	0.2086	-1.53	0.1262	
smokingUnkown:nsurgwounds	-0.6790	0.2046	-3.32	0.0009	***
genderM:exp.muscleYes	1.2438	0.6977	1.78	0.0746	
genderM:diab.type2	1.5734	1.1657	1.35	0.1771	
genderM:nsurgwounds	-0.3667	0.1597	-2.30	0.0216	*
exp.muscleYes:depth	-0.1107	0.0669	-1.66	0.0979	
dorsal.plantarPlantar:depth	-0.1780	0.0778	-2.29	0.0222	*
dorsal.plantarUnknown:depth	-0.0935	0.0753	-1.24	0.2146	
diab.type2:nsurgwounds	-0.8608	0.5592	-1.54	0.1237	

Table 3.3: Estimated coefficients of covariates using logistic regression. "*" in last column is significance code following the rule that "***" means p-value is in (0,0.001) and "*" means that p-value is in (0.01,0.05)

We plot distribution of propensity score as follow.

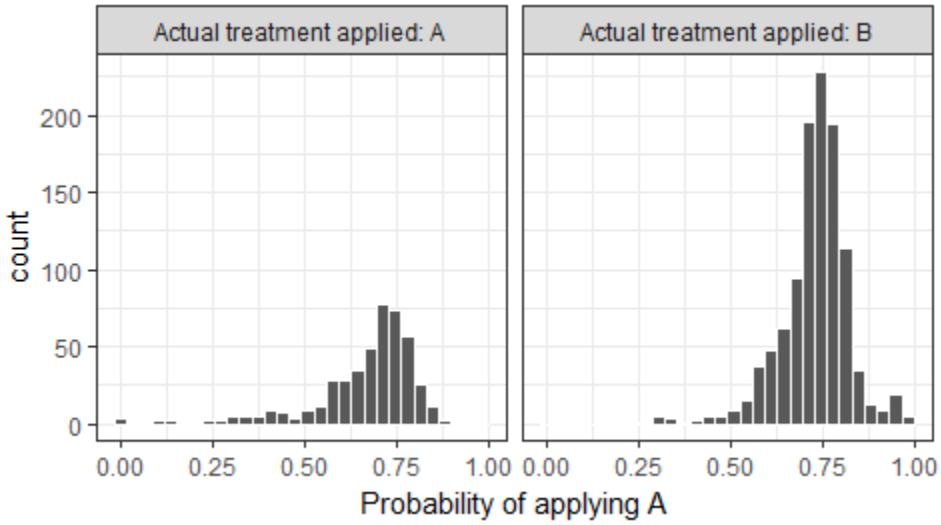

Figure 3.1: Distribution of estimated propensity score using logistic regression. The x-axis shows the propensity scores of two groups and y-axis is the frequency of the propensity score.

Treatment	Min	1st Qu	Median	Mean	3rd Qu	Max	Std
A	0.0002	0.6158	0.7056	0.6656	0.7508	0.9445	0.1411171
B	0.1647	0.6890	0.7381	0.7272	0.7797	0.9813	0.09324231

Table 3.4: Summary of propensity score estimation using logistic regression. It shows the minimum, 1st quartile, median, mean, 3rd quartile, maximum and standard deviation of the propensity scores of two treatment groups.

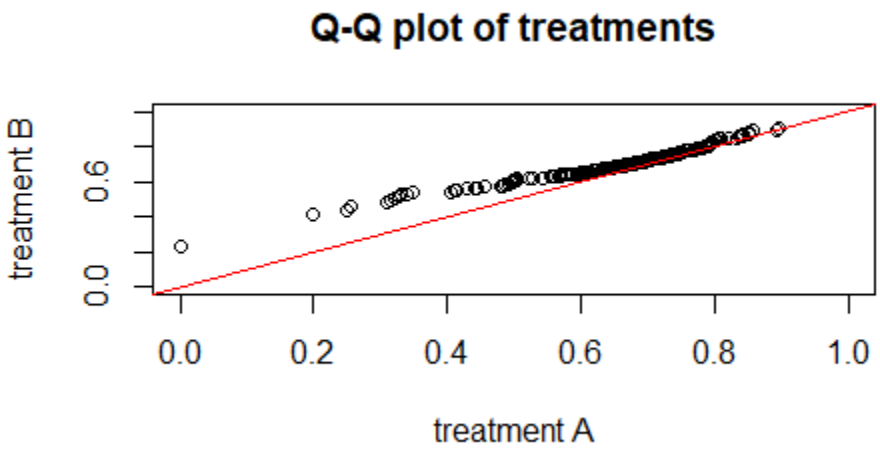

Figure 3.2: Quantile-quantile plot of two treatments

Figure 3.1 shows the estimated propensity scores for given treatment and Table 3.4 summarized them. From them, we can see that the treated group has a larger variance of propensity score than control group. In addition, the ranges of propensity scores of patients from each group are similar, so that we believe propensity scores of patients from both group share common support. We employ an unpaired two samples Wilcoxon test to the estimated propensity scores of patients from different groups. The p-value of the test is quiet small, illustrating that the propensity scores of two treatment groups have different means.

Figure 3.2 illustrates the comparison of the distributions of propensity scores of patients in both groups in terms of quantiles. It suggests that the shapes of the distribution of estimated propensity scores of two treatment groups are somewhat dissimilar.

3.3.2 Generalized Boosted Modeling

Using the *gbm* package in R, a generalized boosted model was fitted with bernoulli loss function. 100 iterations were performed and there are 13 predictors of which 12 had non-zero influence. Table 3.5, the columns "rel.inf" shows the relative influence values of 13 pretreatment covariates, which are calculated using Equation 2.3. From the table, we can see that the width of the wound, which has the highest relative influence, seems to play a key role in influencing treatment assignment. In addition, the depth, area, length of the wound and patient's gender are also somewhat influential to treatment assignment. However, influences do not provide any explanation about how the covariate actually affects the response.

var	rel.inf	var	rel.inf
width	22.11	exp.muscle	4.44
depth	16.82	nsurgwounds	2.53
area	15.64	dorsal.plantar	1.87
gender	14.26	diab.type	0.19
length	10.18	exp.bone	0.16
exposed	6.06	exp.tendon	0.00
smoking	5.74		

Table 3.5: Relative influence calculated by GBM

Figure 3.3 shows the distribution of estimated propensity score using GBM. It seems that these estimated values had narrow range than those estimated using logistic regression. Besides, the shape of the distribution of propensity score in Figure 3.3 is dissimilar to that in Figure 3.1

In Table 3.6, we summarize these estimated propensity score using GBM. It clearly describes that the range of the propensity scores of both groups are smaller than that using logistics regression. In addition, the standard deviations of both groups are relatively small.

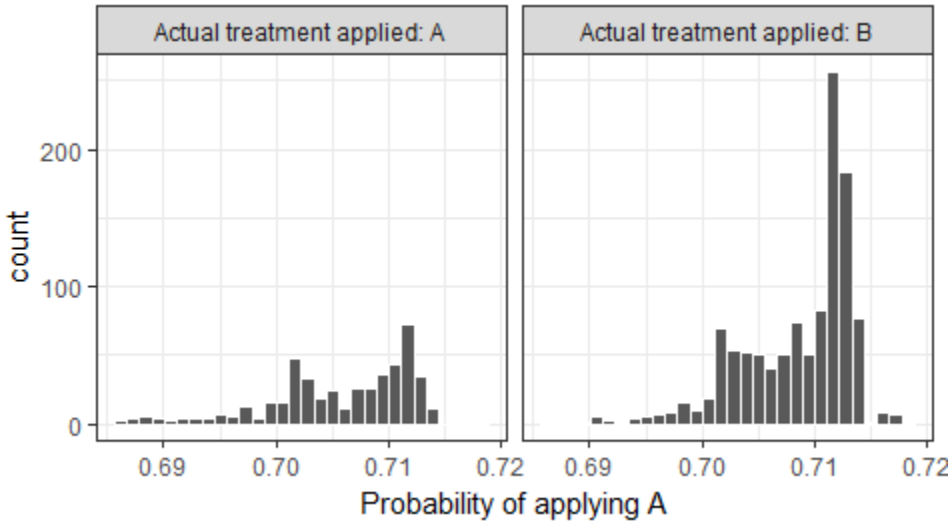

Figure 3.3: Distribution of estimated propensity score using GBM

Employing unpaired two sample Wilcoxon test to these estimated propensity scores, we get p-value < 2.2e-16. Thus, we draw the same conclusion as in Section 3.3.1 that the propensity scores of two treatment groups have different means.

Treatment	Min	1st Qu	Median	Mean	3rd Qu	Max	Std
A	0.6831	0.7022	0.7074	0.7058	0.7109	0.7169	0.0062
B	0.6876	0.7051	0.7106	0.7085	0.7117	0.7165	0.0047

Table 3.6: Summary of propensity score estimation using logistic regression

3.4 Matching and Covariate Balance

Using R package *MatchIt*, we use optimal method for matching patients. And we get the matched pairs as Table 3.7. To illustrate how propensity score matching improves covariate balance, we summarize the covariate imbalance before and after matching of two collections of matched samples. One is matched based on propensity score estimation using logistic regression and the other is matched using BGM estimation.

Sample sizes:		
	Control	Treated
All	1098	450
Matched	450	450
Unmatched	648	0
Discard	0	0

Table 3.7: Matching sample size

Covariates	M_{Xt}	S_{Xt}	M_{Xp}	S_{Xp}	M_{Xc}	S_{Xc}	d_X	d_{X_M}
Smoking Current	0.129	0.335	0.172	0.378	0.111	0.315	0.121	0.055
Smoking Former	0.304	0.461	0.297	0.457	0.313	0.464	0.016	0.019
Smoking Never	0.222	0.416	0.236	0.425	0.220	0.415	0.032	0.005
Smoking Unknown	0.344	0.476	0.295	0.456	0.356	0.479	0.106	0.023
Gender Female	0.244	0.430	0.290	0.454	0.209	0.407	0.102	0.085
Gender Male	0.756	0.430	0.710	0.454	0.791	0.407	0.102	0.085
Exp.tendon No	0.929	0.257	0.957	0.203	0.936	0.246	0.122	0.026
Exp.tendon Yes	0.071	0.257	0.043	0.203	0.064	0.246	0.122	0.026
Exp.muscle No	0.931	0.254	0.971	0.168	0.940	0.238	0.185	0.036
Exp.muscle Yes	0.069	0.254	0.029	0.168	0.060	0.238	0.185	0.036
Exp.bone No	0.924	0.265	0.951	0.216	0.927	0.261	0.109	0.008
Exp.bone Yes	0.076	0.265	0.049	0.216	0.073	0.261	0.109	0.008
Exposed No	0.829	0.377	0.899	0.302	0.844	0.363	0.205	0.042
Exposed Yes	0.171	0.377	0.101	0.302	0.156	0.363	0.205	0.042
Dorsal	0.076	0.265	0.053	0.224	0.062	0.242	0.093	0.053
Plantar	0.367	0.482	0.433	0.496	0.364	0.482	0.135	0.005
Unknown	0.558	0.497	0.515	0.500	0.573	0.495	0.087	0.031
Diab.type 1	0.024	0.155	0.048	0.214	0.036	0.185	0.127	0.065
Diab.type 2	0.976	0.155	0.952	0.214	0.964	0.185	0.127	0.065
Length	18.916	20.994	15.674	15.979	18.219	16.592	0.174	0.037
Width	15.176	13.445	12.911	12.715	15.328	15.195	0.173	0.011
Area	4.713	10.672	3.355	8.439	4.359	8.617	0.141	0.036
Depth	3.581	4.637	2.849	2.934	3.317	3.705	0.189	0.063
Nsurgwounds	1.644	1.046	1.566	0.908	1.622	0.981	0.080	0.022

Table 3.8: ASAM for each covariates before and after matching based on estimated propensity score using logistic regression

Table 3.8 and Table 3.9 show the means of treated group M_{Xt}, entire control group M_{Xp} and matched control group M_{Xc} as well as the standard deviations of of treated group S_{Xt}, entire control group S_{Xp} and matched control group S_{Xc}. Then, ASAMs before matching d_X and after matching d_{X_M} are calculated by Equation 2.6 and Equation 2.7 respectively, which are used to measure covariate imbalance. d_X and d_{X_M} approaching to 0 is an indicator of good balance.

In Table 3.8, there are several covariates that have a much smaller d_{X_M} after matching. For instance, the dummy variable "Plantar" has a $d_X = 0.135$ before matching and it decreases to 0.005 after matching; and the continuous variable "width" has a decrease of imbalance from 0.173 to 0.011. Although, the dummy variable "Smoking Former" has an increase of imbalance from 0.016 to 0.019, 0.019 is also enough small to be regarded as balanced. In conclusion, we suggest the matching improves covariate balance.

In Table 3.9, almost every variable has a decrease of imbalance. However, compared with Table 3.8, these decreases don't imply a great improvement of covariate balance. For instance, all d_{X_M} in Table 3.8 are smaller than 0.1, while some of d_{X_M} in Table 3.9 are still larger than 0.1, such as, "Exposed No", "Exposed Yes" and "Length". Actually, most of the covariates have a smaller d_{X_M} in Table 3.8, which means the logistic regression model performs better in terms of covariate balance in this case. Thus, we apply postmatching analysis to matched data based on logistic regression estimation.

Covariates	M_{Xt}	S_{Xt}	M_{Xp}	S_{Xp}	M_{Xc}	S_{Xc}	d_X	d_{X_M}
Smoking Current	0.13	0.34	0.17	0.38	0.16	0.37	0.12	0.10
Smoking Former	0.30	0.46	0.30	0.46	0.29	0.46	0.02	0.02
Smoking Never	0.22	0.42	0.24	0.42	0.23	0.42	0.03	0.03
Smoking Unknown	0.34	0.47	0.29	0.46	0.31	0.46	0.11	0.07
Gender Female	0.24	0.43	0.29	0.45	0.28	0.45	0.10	0.08
Gender Male	0.76	0.43	0.71	0.45	0.72	0.45	0.10	0.08
Exp.tendon No	0.93	0.26	0.96	0.20	0.95	0.21	0.12	0.10
Exp.tendon Yes	0.07	0.26	0.04	0.20	0.05	0.21	0.12	0.10
Exp.muscle No	0.93	0.25	0.97	0.17	0.97	0.18	0.19	0.16
Exp.muscle Yes	0.07	0.25	0.03	0.17	0.03	0.18	0.19	0.16
Exp.bone No	0.92	0.26	0.95	0.22	0.95	0.22	0.11	0.10
Exp.bone Yes	0.08	0.26	0.05	0.22	0.05	0.22	0.11	0.10
Exposed No	0.83	0.38	0.90	0.30	0.89	0.31	0.21	0.17
Exposed Yes	0.17	0.38	0.10	0.30	0.11	0.31	0.21	0.17
Dorsal	0.08	0.26	0.05	0.22	0.06	0.23	0.09	0.08
Plantar	0.37	0.48	0.43	0.49	0.42	0.49	0.14	0.11
Unknown	0.56	0.50	0.52	0.50	0.52	0.50	0.09	0.07
Diab.type 1	0.02	0.15	0.05	0.21	0.04	0.20	0.13	0.11
Diab.type 2	0.98	0.15	0.95	0.21	0.96	0.20	0.13	0.11
Length	18.92	20.99	15.67	15.98	15.97	16.12	0.17	0.16
Width	15.18	13.45	12.91	12.71	13.23	12.51	0.17	0.15
Area	4.71	10.67	3.35	8.44	3.43	8.32	0.14	0.13
Depth	3.58	4.64	2.85	2.93	2.90	3.08	0.19	0.17
Nsurgwounds	1.64	1.05	1.57	0.91	1.57	0.92	0.08	0.07

Table 3.9: ASAM for each covariates before and after matching based on estimated propensity score using GBM

Chapter 4

Results and Discussion

To estimate treatment effect, we both apply regression adjustment based on matched samples and approximate it by the difference of the averages of outcome variables in treated and control groups. Besides, a doubly robust estimator is also used to draw a safe conclusion.

4.1 Regression adjustment

When we apply regression adjustment, we found that the ATE of whether the wound is closed by 12 weeks, which is $\hat{\alpha}$ in regression model equals to -0.0149. However, the p-value is larger than 0.05. Then we fail to reject the null hypothesis that there is a difference between the treatment effects.

4.2 Difference of Mean of Outcome Variables

When we consider whether the wound is closed by 12 weeks under either treatment, the matched data above shows that, the probability of closure for treatment 1 is 57.5% while for treatment 0 that is 57.8%. p-value for two samples t-test is 0.9394 which is much larger than 0.05. Hence we cannot suggest that there is a difference between the treatment effect of two

treatments.

If we care whether the wound will be closed by 24 weeks, the probability for closure within 24 weeks is 63.5% for treatment 1 and alternatively 62.9% for treatment 0. Similarly, the p-value, 0.8762, is beyond the critical threshold. Thus, we still cast doubt on that there is a significant difference.

4.3 Doubly Robust

Using the matched data, fitting general linear models is another approach to estimate the treatment effect. Once we assume that the outcome binary variable and covariates fit a general linear regression model, logistic model for instance, we predict the fitted value of outcome for both treatments on the covariates data we have.

And then, we compare the mean of two groups of fitted values. The t-test shows that the p-value is 0.098, which is also larger than 0.05. It reaches the same conclusion as we directly compare the observed outcomes.

All three methods show that we cannot conclude that there is a difference between treatment effects.

4.4 Discussion

The paper completed a case of analysis of treatment effects. But there are still several problems that can be taken into account in the future.

First, in the application in this paper, logistic regression modeling performs better than generalized boosted modeling in terms of estimating propensity score. But GBM has several

advantages. For instance, it can capture complex and nonlinear relationships between covariates and outcomes without overfitting the data. So, it comes to a thinking about that when should one use GBM. Besides, in this application of GBM, we do not take interaction terms into consideration, which may be associated with treatment assignment.

Additionally, propensity score can be used to subclassification and weighting for treated and control samples. There are several methods concerning causal inference using propensity score weighting estimators. In some case, one should use weighting estimator instead of matching estimators.

Third, in our application, we selected potential covariates following the instruction of Hirano, Imbens, and Ridder (2003). The methods can be spread to a sensitivity analysis that it tests how sensitive the estimated treatment effect is to diverse settings of logistic regression model and postmatching analysis.

References

Abadie, A., & Imbens, G. W. (2011). Bias-Corrected Matching Estimators for Average Treatment Effects. *J. Bus. Econ. Stat.*, *29*(1), 1–11. Retrieved from https://doi.org/10.1198/jbes.2009.07333 doi: 10.1198/jbes.2009.07333

Austin, P. C. (2011, mar). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharm. Stat.*, *10*(2), 150–161. Retrieved from http://doi.wiley.com/10.1002/pst.433 doi: 10.1002/pst.433

Baser, O. (2006). Too Much Ado about Propensity Score Models? Comparing Methods of Propensity Score Matching. *Value Heal.*, *9*(6), 377 – 385. Retrieved from http://www.sciencedirect.com/science/article/pii/S109830151063152X doi: https://doi.org/10.1111/j.1524-4733.2006.00130.x

Breiman, L. (2001). Statistical Modeling: The Two Cultures. *Stat. Sci.*, *16*(3), 199–231. doi: 10.1214/ss/1009213726

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. CRC press.

Caliendo, M., & Kopeinig, S. (2008, feb). Some Practical Guidance for the Implementation of Propensity Score Matching. *J. Econ. Surv.*, *22*(1), 31–72. Retrieved from http://doi.wiley.com/10.1111/j.1467-6419.2007.00527.x doi: 10.1111/j.1467-6419.2007.00527.x

Chan, K. C. G., Yam, S. C. P., & Zhang, Z. (2016, jun). Globally efficient non-parametric inference of average treatment effects by empirical balancing calibration weighting. *J. R. Stat. Soc. Ser. B (Statistical Methodol.*, *78*(3), 673–700. Retrieved from http://doi.wiley.com/10.1111/rssb.12129 doi: 10.1111/rssb.12129

Chang, C.-C. H. (2017, may). Applications of the propensity score weighting method in psychogeriatric research: correcting selection bias and adjusting for confounders. *Int. Psychogeriatrics*, *29*(05), 703–706. Retrieved from https://www.cambridge.org/core/product/identifier/S1041610216002490/type/journal{_}article doi: 10.1017/S1041610216002490

Cochran, W. G., & Rubin, D. B. (1973). Controlling Bias in Observational Studies: A

Review. *Sankhyā Indian J. Stat. Ser. A*, *35*(4), 417–446. Retrieved from http://www.jstor.org/stable/25049893

DB, R. (1997, oct). EStimating causal effects from large data sets using propensity scores. *Ann. Intern. Med.*, *127*(8_Part_2), 757–763. Retrieved from http://dx.doi.org/10.7326/0003-4819-127-8{_}Part{_}2-199710151-00064

Dehejia, R. H., & Wahba, S. (2002). Propensity Score-Matching Methods for Nonexperimental Causal Studies. *Rev. Econ. Stat.*, *84*(1), 151–161. Retrieved from https://doi.org/10.1162/003465302317331982 doi: 10.1162/003465302317331982

Elith, J., Leathwick, J. R., & Hastie, T. (2008, jul). A working guide to boosted regression trees. *J. Anim. Ecol.*, *77*(4), 802–813. Retrieved from http://doi.wiley.com/10.1111/j.1365-2656.2008.01390.x doi: 10.1111/j.1365-2656.2008.01390.x

Fife, C. E., Horn, S. D., Smout, R. J., Barrett, R. S., & Thomson, B. (2016). A Predictive Model for Diabetic Foot Ulcer Outcome: The Wound Healing Index. *Adv. Wound Care*, *5*(7), 279–287. doi: 10.1089/wound.2015.0668

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Ann. Stat.*, 1189—-1232.

Friedman, J. H. (2002). Stochastic gradient boosting. *Comput. Stat. Data Anal.*, *38*(4), 367–378. Retrieved from http://www.sciencedirect.com/science/article/pii/S0167947301000652 doi: https://doi.org/10.1016/S0167-9473(01)00065-2

Funk, M. J., Westreich, D., Wiesen, C., Stürmer, T., Brookhart, M. A., & Davidian, M. (2011, apr). Doubly Robust Estimation of Causal Effects. *Am. J. Epidemiol.*, *173*(7), 761–767. Retrieved from http://dx.doi.org/10.1093/aje/kwq439

Garrido, M. M., Kelley, A. S., Paris, J., Roza, K., Meier, D. E., Morrison, R. S., & Aldridge, M. D. (2014, oct). Methods for Constructing and Assessing Propensity Scores. *Health Serv. Res.*, *49*(5), 1701–1720. Retrieved from http://doi.wiley.com/10.1111/1475-6773.12182 doi: 10.1111/1475-6773.12182

Guo, S., & Fraser, M. W. (2014). *Propensity Score Analysis* (2nd ed.). Sage.

Harder, V. S., Stuart, E. A., & Anthony, J. C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychol. Methods*, *15*(3), 234–249. doi: 10.1037/a0019623

Haviland, A., Nagin, D. S., & Rosenbaum, P. R. (2007). Combining propensity score matching and group-based trajectory analysis in an observational study. *Psychol. Methods*, *95*(450), 573–585. doi: 10.1037/1082-989X.12.3.247

Helmreich, J. E., & Pruzek, R. M. (2009). PSAgraphics: An R Package to Support Propensity Score Analysis. *J. Stat. Softw.*, *29*(6), 1–23. Retrieved from http://www.jstatsoft.org/

Hirano, K., Imbens, G. W., & Ridder, G. (2003, jul). Efficient Estimation of Average

Treatment Effects Using the Estimated Propensity Score. *Econometrica*, *71*(4), 1161–1189. Retrieved from http://doi.wiley.com/10.1111/1468-0262.00442 doi: 10.1111/1468-0262.00442

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt : Nonparametric Preprocessing for Parametric Causal Inference. *J. Stat. Softw.*, *42*(8), 1–28. doi: 10.18637/jss.v042.i08

Ho, D. E., Imai, K., King, G., Stuart, E. A., Abadie, A., Beck, N., ... Rubin, D. (2007). Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Polit. Anal.*, *15*, 199–236. doi: 10.1093/pan/mpl013

Hoffstad, O., Mitra, N., Walsh, J., & Margolis, D. J. (2015). Diabetes, Lower-Extremity Amputation, and Death. *Diabetes Care*, *38*(10), 1852–1857. Retrieved from http://care.diabetesjournals.org/content/38/10/1852 doi: 10.2337/dc15-0536

Horn, S. D., Fife, C. E., Smout, R. J., Barrett, R. S., & Thomson, B. (2013, nov). Development of a wound healing index for patients with chronic wounds. *Wound Repair Regen.*, *21*(6), 823–832. Retrieved from http://doi.wiley.com/10.1111/wrr.12107 doi: 10.1111/wrr.12107

Imbens, G. W. (2004). Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review. *Rev. Econ. Stat.*, *86*(1), 4–29. Retrieved from https://doi.org/10.1162/003465304323023651 doi: 10.1162/003465304323023651

Keller, B., & Tipton, E. (2016, jun). Propensity Score Analysis in R: A Software Review. *J. Educ. Behav. Stat.*, *41*(3), 326–348. Retrieved from http://journals.sagepub.com/doi/abs/10.3102/1076998616631744 doi: 10.3102/1076998616631744

Keller, B. S. B., Kim, J.-S., & Steiner, P. M. (2010). Propensity Score Estimation with Data Mining Techniques: Alternatives to Logistic Regression. *J Clin Epidemiol.*, *6*(38), 826–833.

Kim, M. S., Koh, I. J., Choi, Y. J., Lee, J. Y., & In, Y. (2017). Differences in Patient-Reported Outcomes Between Unicompartmental and Total Knee Arthroplasties: A Propensity Score-Matched Analysis. *J. Arthroplasty*, *32*(5), 1453 – 1459. Retrieved from http://www.sciencedirect.com/science/article/pii/S0883540316308348 doi: https://doi.org/10.1016/j.arth.2016.11.034

King, G., & Nielsen, R. (2015). Why Propensity Scores Should Not Be Used for Matching *.

Kirsner, R. S., Sabolinski, M. L., Parsons, N. B., Skornicki, M., & Marston, W. A. (2015, sep). Comparative effectiveness of a bioengineered living cellular construct vs. a dehydrated human amniotic membrane allograft for the treatment of diabetic foot ulcers in a real world setting. *Wound Repair Regen.*, *23*(5), 737–744. Retrieved from http://doi.wiley.com/10.1111/wrr.12332 doi: 10.1111/wrr.12332

Lee, B. K., Lessler, J., & Stuart, E. A. (2009, feb). Improving Propensity Score Weighting Using Machine Learning. *Stat. Med.*, *29*(3), n/a–n/a. Retrieved from http://doi.wiley.com/10.1002/sim.3782 doi: 10.1002/sim.3782

Linden, A., & Yarnold, P. R. (2017, aug). Using classification tree analysis to generate propensity score weights. *J. Eval. Clin. Pract.*, *23*(4), 703–712. Retrieved from http://doi.wiley.com/10.1111/jep.12744 doi: 10.1111/jep.12744

McCaffrey, D. F., Griffin, B. A., Almirall, D., Slaughter, M. E., Ramchand, R., & Burgette, L. F. (2013, aug). A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Stat. Med.*, *32*(19), 3388–3414. Retrieved from http://doi.wiley.com/10.1002/sim.5753 doi: 10.1002/sim.5753

McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychol. Methods*, *9*(4), 403–425.

Mease, D., Wyner, A. J., & Buja, A. (2007). Boosted Classification Trees and Class Probability/Quantile Estimation. *J. Mach. Learn. Res.*, *8*, 409–439.

Pirracchio, R., Petersen, M. L., & van der Laan, M. (2015, jan). Improving Propensity Score Estimators' Robustness to Model Misspecification Using Super Learner. *Am. J. Epidemiol.*, *181*(2), 108–119. Retrieved from http://dx.doi.org/10.1093/aje/kwu253

Ridgeway, G. (1999). The state of boosting. *Comput. Sci. Stat.*, 172—-181.

Rosenbaum, P. R. (2002). Covariance Adjustment in Randomized Experiments and Observational Studies. *Stat. Sci.*, *08*(3), 286–327. doi: 10.1214/ss/1042727942

Rosenbaum, P. R., & Rubin, D. B. (1983, apr). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. Retrieved from http://dx.doi.org/10.1093/biomet/70.1.41

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. *J. Am. Stat. Assoc.*, *79*(387), 516–524. Retrieved from http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1984.10478078 doi: 10.1080/01621459.1984.10478078

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. *Am. Stat.*, *39*(1), 33–38. Retrieved from http://amstat.tandfonline.com/doi/abs/10.1080/00031305.1985.10479383 doi: 10.1080/00031305.1985.10479383

Rubin, D. B. (1974, dec). Multivariate Matching Methods That are Equal Percent Bias Reducing, II: Maximums on Bias Reduction for Fixed Sample Sizes. *ETS Res. Bull. Ser.*, *1974*(2), i–26. Retrieved from http://doi.wiley.com/10.1002/j.2333-8504.1974.tb00866.x doi: 10.1002/j.2333-8504.1974.tb00866.x

Rubin, D. B. (1977, mar). Assignment to Treatment Group on the Basis of a Covariate. *J. Educ. Stat.*, *2*(1), 1–26. Retrieved from http://journals.sagepub.com/doi/abs/10.3102/10769986002001001 doi: 10.3102/10769986002001001

Rubin, D. B. (1978a, dec). Bias Reduction Using Mahalanobis-Metric Matching. *ETS Res. Bull. Ser.*, *1978*(2), i–10. Retrieved from http://doi.wiley.com/10.1002/j.2333-8504

.1978.tb01164.x doi: 10.1002/j.2333-8504.1978.tb01164.x

Rubin, D. B. (1978b, dec). USING MULTIVARIATE MATCHED SAMPLING AND REGRESSION ADJUSTMENT TO CONTROL BIAS IN OBSERVATIONAL STUDIES. *ETS Res. Bull. Ser.*, *1978*(2), i–33. Retrieved from http://doi.wiley.com/10.1002/j.2333-8504.1978.tb01163.x doi: 10.1002/j.2333-8504.1978.tb01163.x

Rubin, D. B., & Thomas, N. (2000). Combining Propensity Score Matching with Additional Adjustments for Prognostic Covariates. *J. Am. Stat. Assoc.*, *95*(450), 573–585. Retrieved from http://amstat.tandfonline.com/doi/abs/10.1080/01621459.2000.10474233 doi: 10.1080/01621459.2000.10474233

Rudolph, K. E., Colson, K. E., Stuart, E. A., & Ahern, J. (2016, nov). Optimally combining propensity score subclasses. *Stat. Med.*, *35*(27), 4937–4947. Retrieved from http://doi.wiley.com/10.1002/sim.7046 doi: 10.1002/sim.7046

Uysal, S. D. (2015, aug). Doubly Robust Estimation of Causal Effects with Multivalued Treatments: An Application to the Returns to Schooling. *J. Appl. Econom.*, *30*(5), 763–786. Retrieved from http://doi.wiley.com/10.1002/jae.2386 doi: 10.1002/jae.2386

with contributions from others, G. R. (2017). gbm: Generalized boosted regression models [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=gbm (R package version 2.1.3)

Appendix A

Appendix: R code

```
library(MASS)
library(MatchIt)
library(dplyr)
library(ggplot2)
library(gbm)
library(locfit)
library(twang)
library(nnet)

treatment <- dfu.all$treatment

######list out the categorical variables and find possible influential ones
covariates.list <- c(#"body.part",
  "smoking","gender","exp.tendon",
  "exp.ligament","exp.muscle","exp.joint","exp.bone","exposed",
  "dorsal.plantar","proximal.distal","diab.type",
  "age.cat")

#################fitting contingency tables######################
for(i in covariates.list)
  {
    txt <- paste("x <- as.factor(dfu.all$",i,")",sep = "")
    eval(parse(text=txt))
    table.i <- table(treatment,x)
    f <- fisher.test(table.i)
    if(f$p.value < 0.2 & !is.na(f$p.value))  {print(i)}
  }
```

```
df.all <- data.frame(
  treatment=dfu.all$treatment,
  smoking=dfu.all$smoking,
  gender=dfu.all$gender,
  exp.tendon=dfu.all$exp.tendon,
  exp.muscle=dfu.all$exp.muscle,
  exp.bone=dfu.all$exp.bone,
  exposed=dfu.all$exposed,
  dorsal.plantar=dfu.all$dorsal.plantar,
  diab.type=dfu.all$diab.type
)

n <- dim(dfu.all)[1]

continuous.list <- c("age","bmi","length","width","area","depth","ndfu",

"nwounds","nsurgwounds")

####################Applying logistic regression#######################
for(i in continuous.list)
{
  txt <- paste("x␣<-␣as.numeric(dfu.all$",i,")",sep = "")
  eval(parse(text=txt))
  model <- glm(treatment~x,family = binomial(link = "logit"))
  p.value <- summary(model)$coef[2,4]
  if(p.value < 0.2 & !is.na(p.value))  {print(i)}
}

df.all$length <- as.numeric(dfu.all$length)
df.all$width <- as.numeric(dfu.all$width)
df.all$area <- as.numeric(dfu.all$area)
df.all$depth <- as.numeric(dfu.all$depth)
df.all$nsurgwounds <- as.numeric(dfu.all$nsurgwounds)


################Estimating Propensity score################

#########Logistic Regression

####Variable Selection
```

```
df <- na.omit(df.all)
fit1 <- glm(treatment~.,data = df,family=binomial(link = "logit"))

fit <- stepAIC(fit1,direction = "both")

##treatment ~ smoking + gender + exp.muscle +
##               dorsal.plantar + diab.type + width + depth + nsurgwounds

####Estimate the propensity score
fit <- glm(formula = treatment ~ smoking + gender + exp.muscle +
              dorsal.plantar + diab.type + width + depth + nsurgwounds,
             family = binomial(link = "logit"), data = df)
prs_df <- data.frame(pr_score = predict(fit, type = "response"),
                     treatment = fit$model$treatment)
head(prs_df,n=10)

#####summarize pr_score
summary(prs_df$pr_score[prs_df$treatment=="A"])
sd(prs_df$pr_score[prs_df$treatment=="A"])
summary(prs_df$pr_score[prs_df$treatment=="B"])
sd(prs_df$pr_score[prs_df$treatment=="B"])

qqplot(prs_df$pr_score[prs_df$treatment=="A"],
       prs_df$pr_score[prs_df$treatment=="B"],
       xlim = c(0,1), ylim = c(0,1),
       xlab = "treatment A", ylab = "treatment B",
       main = "Q-Q plot of treatments")
abline(0,1,col=2)

###  Examine the region of common support
labs <- paste("Actual treatment applied:", c("B", "A"))
prs_df %>%
  mutate(treatment = ifelse(treatment == "B", labs[1], labs[2])) %>%
  ggplot(aes(x = pr_score)) +
  geom_histogram(color = "white") +
  facet_wrap(~ treatment) +
  xlab("Probability of applying A") +
  theme_bw()

#########     Generalized Boosted Modeling      ###########
```

```
df.all$treatment <- as.numeric(ifelse(df.all$treatment=="A",1,0))

gbmodel <- gbm(treatment~.,data = df.all,
               distribution = "bernoulli",interaction.depth = 3,
             shrinkage = 0.001,cv.folds = 5)

gbm.perf(gbmodel,plot.it = TRUE)
summary(gbmodel)

####Estimate the propensity score
prs_gbm <- data.frame(pr_score = 1-expit(gbmodel$fit),
                      treatment=df.all$treatment)
head(prs_gbm,n=10)

#####summarize pr_score
summary(prs_gbm$pr_score[prs_df$treatment=="A"])
sd(prs_gbm$pr_score[prs_df$treatment=="A"])
summary(prs_gbm$pr_score[prs_df$treatment=="B"])
sd(prs_gbm$pr_score[prs_df$treatment=="B"])

###  Examine the region of common support
labs <- paste("Actual␣treatment␣applied:", c("B", "A"))
prs_gbm %>%
  mutate(treatment = ifelse(treatment == "0", labs[1], labs[2])) %>%
  ggplot(aes(x = pr_score)) +
  geom_histogram(color = "white") +
  facet_wrap(~ treatment) +
  xlab("Probability␣of␣applying␣A") +
  theme_bw()

#######################Matching#########################
df <- data.frame(df.all,closed.12.weeks = dfu.all$closed.12.weeks)
df_cov <- names(df.all)[-1]
df_nomiss <- df %>%
  select(closed.12.weeks, treatment, one_of(df_cov)) %>%
  na.omit()
#m <- matchit(formula(fit), method = "nearest", data = df_nomiss)
m <- matchit(formula(fit), method = "optimal", data = df_nomiss)
m
matched <- match.data(m)
```

```
####################Covariate Imbalance
df.all <- na.omit(df.all)
p <- dim(df.all)[2]

df <- data.frame(treatment = df.all$treatment)
df_m <- data.frame(close.12.weeks = matched$closed.12.weeks,
                   treatment = matched$treatment)

for(i in 2:p){
  if(class(df.all[,i])=="numeric")
  {
    df <- data.frame(df,df.all[,i])
    df_m <- data.frame(df_m,matched[,i+1])
  }else{
    temp <- class.ind(df.all[,i])
    temp_m <- class.ind(matched[,i+1])
    df <- data.frame(df,temp)
    df_m <- data.frame(df_m,temp_m)
  }
}

q <- dim(df)[2]   ###     # of "covariate"
blc <- matrix(nrow = q-1,ncol = 8)
for(i in 2:q){
  blc[i-1,1] <- mean(df[which(df$treatment=="1"),i])
  blc[i-1,2] <- sd(df[which(df$treatment=="1"),i])
  blc[i-1,3] <- mean(df[which(df$treatment=="0"),i])
  blc[i-1,4] <- sd(df[which(df$treatment=="0"),i])
  s <-sqrt((blc[i-1,2]^2+blc[i-1,4]^2)/2)
  blc[i-1,5] <- mean(df_m[which(matched$treatment=="0"),i+1])
  blc[i-1,6] <- sd(df_m[which(matched$treatment=="0"),i+1])
  s2 <- sqrt((blc[i-1,2]^2+blc[i-1,6]^2)/2)
  blc[i-1,7] <- abs(blc[i-1,1]-blc[i-1,3])/s
  blc[i-1,8] <- abs(blc[i-1,1]-blc[i-1,5])/s2
}

####################GBM analysis#############################
psmodel <- ps(treatment~.,data = df.all,interaction.depth = 2,
              estimand = "ATE")

plot.ps(psmodel)
```

```
m.balance <- bal.table(psmodel)
m.balance
q <- dim(df)[2]-1   ###     # of "covariate"
blc <- matrix(nrow = q,ncol = 8)
for(i in 1:q){
  blc[i,1:4] <- as.numeric(m.balance$unw[i,c("tx.mn","tx.sd","ct.mn","ct.sd")])
  s <-sqrt((blc[i,2]^2+blc[i,4]^2)/2)
  blc[i,5:6] <- as.numeric(m.balance$es.mean.ATE[i,3:4])
  s2 <- sqrt((blc[i,2]^2+blc[i,6]^2)/2)
  blc[i,7] <- abs(blc[i,1]-blc[i,3])/s
  blc[i,8] <- abs(blc[i,1]-blc[i,5])/s2
}
xtable(blc)

#################3Regression Adjustment####################
df_m_tr <- df_m[which(df_m$treatment==1),]
df_m_ct <- df_m[which(df_m$treatment==0),]
data <- df_m_tr-df_m_ct
fit <- lm(close.12.weeks~.,data = data)
summary(fit)

###############Average Over Outcome Variables#############
with(matched,t.test(closed.12.weeks~treatment))
with(matched,t.test(closed.24.weeks~treatment))
```

Chenqi Fu

500 W University Pkwy APT6G, Baltimore, MD 21210

443-703-8028

cfu15@jhu.edu

EDUCATION

Johns Hopkins Whiting School of Engineering — Baltimore, MD
Master of Science and Engineering in Applied Mathematics and Statistics — *09/2016-present*

Fudan University — Shanghai,China
Bachelor of Science in Mathematics — *09/2012-07/2016*
Exchanged to Washington University in St.Louis — *Summer 2015*

RESEARCH EXPERIENCE

Propensity Scoring Comparison of Wound Care Treatments — 06/2017-present
Johns Hopkins University

- Extracted information of subjects from database to form a final collection of covariates and outcome variables
- Calculated descriptive statistics among the samples
- Fitted logistic regression models and decision tree to model propensity scores
- Applied matching method to adjust data and applied hypothesis test
- Used doubly robust method to estimate treatment effect
- Contributed to report writing

Evolution strategies for meta-learning and network learning — 03/2017-06/2017
*Machine Learning,Johns Hopkins University**

- Programmed codes in Python using evolution strategy for meta-learner
- Trained agents on LunarLander-v2 tasks from OpenAI
- Contributed to final presentation in class and report writing

Thesis: China Household Finance Survey — 09/2015-01/2016
Fudan University

- Extracted data concerning family education expense from database
- Categorized covariates and applied contingency tables
- Applied multiple imputation for missing data and fitted logistics regression model and lasso model
- Contributed to final presentation and thesis writing

Cell Identification from Calcium Imaging Data — 09/2015-01/2016
State Key Laboratory of Brain Science, Fudan University

- Reviewed literatures
- Grouped team members and led discussion

WORK EXPERIENCE

Johns Hopkins Whiting School of Engineering — Baltimore, MD
Teaching Assistant — *01/2017-06/2017*

- Statistical Analysis I by Dr.Fred Torcaso
- Statistical Analysis II by Dr.Avanti Athreya

Johns Hopkins Whiting School of Engineering — Baltimore, MD
Introduction — *Intersession 2017*

– Introduction to R

VONLUNTEER WOKRS

Volunteer Association - Sunflower of Fudan University Shanghai, China
Member *09/2013-09/2014*

– Worked as pre-school facilitator for children of migrant workers every Friday night for a whole term

– Tutored a pupil every Sunday afternoon for a whole term

HONORS

2014 - 2015 Fudan University Outstanding Students Scholarship

2013 - 2015 University Undergraduates Major Scholarship (Twice) 2013 -

2012 - 2013 Xu Zengshou Scholarship (Top 5)

11/2013 The Chinese Mathematics Competition in Shanghai, 2rd place

PROFESSIONAL DEVELOPMENT

Computer Skills: R; Python; C; Matlab; Microsoft Office products (Word, Excel, and Powerpoint); Latex

Language: Full professional proficiency in English; Native proficiency in Chinese